Policy Sciences 13 (1981) 247–267
Elsevier Scientific Publishing Company, Amsterdam – Printed in the Netherlands

The Forecasting Potential of Complex Models

WILLIAM ASCHER
Department of Political Science, The Johns Hopkins University, Baltimore, MD, U.S.A.

ABSTRACT

The nature and use of "complex" models for forecasting and policy simulation are analyzed on theoretical and empirical-performance grounds. The analysis suggests that while the accuracy of complex models in forecasting trends in such fields as economic and energy is, and will remain, undistinguished, complex models' special virtues of preserving counter-intuitive results and representing subsystem interdepedence could be used to better advantage than current practice permits. Suggestions for such improvements, through more diversified model structures, micro-process models in addition to the typical macro models, a mix of mechanically- and judgmentally-operated models, and the modeling of policy response, are reviewed.

1. Introduction

Complex models used in forecasting are expensive, cumbersome, time consuming, and often barely manageable. Defined as a set of two or more explicit propositions that share at least one factor or variable, the complex model's interconnectedness usually gives rise to intricate formulations that can intimidate both the model's creators and the users of its forecasts [1]. Most tellingly, the types of complex models employed in forecasting (econometric models, systems dynamics, etc.) have not been able to justify complex modeling on the basis of the existing record of forecast accuracy.

Yet judging by the volume of research devoted to model development and the money spent on subscriptions to forecast-modeling services, these models are enormously attractive to many forecasters and forecast users. Unless forecasters are completely ignorant of the performance record, or are attracted solely by the promotional advantages of the scientific aura of modeling, they can only be attracted to its potential benefits not yet realized. There is no question that complex models do

0032-2687/81/0000-0000/$02.50

have unique properties as forecasting tools, just as they do for analysis in general. Multiple, interconnected propositions (i.e., outputs of one become inputs of another, or solutions must satisfy several propositions or equations simultaneously) can explicitly and systematically express mutual causation, feedback phenomena, and other intricate relationships. The question is whether these properties can be harnessed to improve prediction and explanation.

To some extent, this potential is revealed by the existing performance record of forecasting models. Yet it may be argued that complex modeling is such a recent development in practical forecasting that its track record is not an adequate basis of evaluation, because the accuracy of many recent forecasts produced by complex models cannot yet be determined, and current models will presumably be further refined and elaborated through efforts to improve them. But even if the record of complex models is an insufficient basis of appraisal, the record of forecasting in general reveals much about the nature of forecasting and the nature of the world that the forecasts address. In addition, we may also consider the logical and structural properties of complex models, and evaluate these properties with respect to the task of forecasting. Thus, if indeed it is premature to judge the forecasting potential of modeling solely on the basis of the record to date, other bases are available. These are: the nature of complex modeling, the nature of forecasting, and the nature of the practical application of modeling to forecasting. After reviewing the performance record of complex models in forecasting, each of these topics will be considered in turn. Finally, some conclusions regarding the optimal use of complex models in forecasting will be drawn from these considerations.

2. The Performance of Complex Models in Forecasting

Complex models have been prominent in projecting only a few relevant policy trends. For many years now, econometric models have been used in short-term economic forecasting, and more recently complex models have been employed for energy demand projections. In population forecasting, however, the impressive-seeming 'models' utilized by the Census Bureau are not complex models at all according to our definition, since they lack connected propositions [2]. They are, in reality, elaborate accounting devices, tracking birth and death rates implied by component- and cohort-specific fertility and mortality rates, but without any of the interactive effects required by our definition of the complex model. Similarly, travel-demand forecasting models, also often formidable in appearance, are nonetheless usually single-equation regression models treating the demand for a particular travel mode strictly as a dependent variable [3]. No matter how many independent variables are included in such an equation, they are not explained or affected by the single dependent variable, but rather projected independently of the models, (i.e., taken exogenously from extrapolations or other models). Consequently, the performance record of complex modeling rests largely on the economic and energy forecasts, although complex modeling could be applied to forecasting in any of the other areas.

In appraising the complex models used in forecasting, there are two questions we may ask of the performance record. The first is how well complex models forecast, whether in absolute terms or in comparison with other forecasting methods. The second, equally important question is whether complex forecasting models, subject to considerable refinement and elaboration over the past two decades, have been improving in terms of their accuracy.

To answer these questions, it is useful to take advantage of the fact that complex models do compete with numerous other techniques as forecasting tools. The forecasting performance of other methods provides not only a comparative basis for evaluating the record of complex modeling, but also a benchmark for establishing the "intrinsic" difficulty of predicting trends of a given period. For example, projecting energy demand trends after 1973 (i.e., predicting the demand in particular target years beyond 1973) has been a more difficult task, both for models and for other forecasting techniques such as extrapolation or judgment, than projecting these trends for target years prior to 1973. (Ascher, 1978, Ch. 5). Thus, it is very useful to regard the average error of "judgmental" forecasts as an indicator of the inherent (rather than method-specific) unpredictability or uncertainty of a specified era. By tracking the level of judgmental error over time (i.e., of forecasts made in years t_1, t_2, t_3, etc., for target years t_{1+N}, t_{2+N}, t_{3+N}, etc., where N is a fixed length of time, say, 5 or 10 years), we can determine whether changes in the accuracy levels of more explicit methods such as modeling can be attributed to changes in predictability. Thus, for example, if the judgmentally-based 5-year petroleum demand forecasts made during the period 1969 to 1973 (for the target years 1974 through 1978) have a higher average error than the judgmental 5-year projections made in, say, 1960–1967 (for the target years 1965 through 1972), then our assessment of less accurate complex-model projections made in 1969–73 for 1974–78 ought to take into account the greater uncertainty of the post-1973 period as indicated by the decline in judgmental accuracy.

2.1. The Record of Economic Forecasting Models

Econometric models for short-term economic forecasting have been evaluated periodically by the model operators themselves (who generally developed econometric forecasting models initially as theoretical experiments, but increasingly as practical commercial enterprises) and more systematically by an ongoing survey sponsored jointly by the American Statistical Association and the National Bureau of Economic Research (ASA-NBER) [4]. This survey provides the most telling appraisal of how well econometric models forecast.

The ASA-NBER survey indicates that econometric models have not, and still do not, forecast quite as well as the judgmental approach relying on no explicit routines.

The most comprehensive ASA-NBER comparison of a large number of economic forecasting sources covered forecasts produced from 1968 through 1973, (V. Su and J. Su, 1975), and found that judgmental forecasting efforts slightly outperformed econometric approaches for quarterly GNP (both nominal and real) and were roughly even in forecasting annual GNP changes.

Although there has been no attempt to repeat so comprehensive a comparison for more recent forecasts, there are some up-to-date comparisons of selected models and non-econometric efforts. The median forecast of the ASA-NBER survey has been used as one benchmark for evaluating the accuracy of prominent modeling operations; the accuracy of forecasts by the Council of Economic Advisors is another non-econometric benchmark. For very short-term forecasts (i.e., one or two quarters), Vincent Su found that for the period 1968–77 the ASA-NBER median forecast still outperforms the Wharton model for most variables, but there is little difference for longer horizons [5]. Victor Zarnowitz found that for 1969–76, the annual forecasts of real GNP and the inflation rate by two modeling operations (Wharton and the University of Michigan model) and the two non-econometric sources do not establish the superiority of one method over another. Zarnowitz concludes: "This is in agreement with earlier findings, which strongly suggests that the search for a consistently superior forecaster is about as promising as the search for the philosophers' stone" (Zarnowitz, 1978). Finally, Stephen McNees found the same two non-econometric sources to be roughly equal in accuracy to several econometric operations (Chase Econometrics and DRI in addition to Wharton and Michigan) in annual forecasting for the 1961–76 period (Mc Nees, 1976, 1977). Thus, there seems to be no dissent from the conclusions that even the most recent econometric modeling has not offered the advantage of greater accuracy over judgment in short-term economic forecasting, despite the refinement and elaboration of the models.

Does this mean, though, that the models per se forecast with the same general level of accuracy as expert opinion? If so, this would be no mean accomplishment. Yet, it is untrue. The prominent econometric modeling services all operate with a considerable amount of judgmental input, introduced either by choosing exogenous variables so that the model's predictions seem plausible to the model operator, or by tinkering with the model specifications or parameters to achieve the same result [6]. Zarnowitz points out that "the genuine ex ante forecasts here considered are all to a large extent 'judgmental' . . ." (Zarnowitz, p. 315).

How do we know that pure models (i.e., the mechanical operation of the models, insulated from judgment) would not do as well or better than the models as they are run today? There are three pieces of evidence. First, the fact that the modelers leave themselves this leeway indicates their appreciation for adding the "polish" of judgment. Second, a valiant attempt by the economist Ray Fair to operate an econometric forecasting model without judgmental adjustments led to a very poor forecasting record; parallel attempts to operate the Wharton model strictly according to its published specifications led to inferior performance than that of the judgmentally modified forecasts actually released by Wharton (Ascher, 1978, Ch. 4 and Klein, 1968). Third, the acutal forecasting errors of the models are lower than the errors produced by operating them with the actual values of exogenous variables once these values are known (Hickman, 1972). This superiority of so-called ex ante predictions over ex post predictions is quite startling; knowing what the actual external reality is like reduces the accuracy of model forecasts. This demonstrates that the convergence of plausible results provided by the forecasters' selection of exogenous values (even if they are, in retrospect, incorrect) is needed to offset errors attributable directly to the models' specifications.

2.2. The Record of Energy Forecasting Models

Appraising the performance record of energy models is complicated by the widespread use of conditional forecasts in this area, as well as by the fact that most energy forecasting models are designed to project long-range trends; their accuracy cannot be evaluated retrospectively as yet. A brief methodological digression may clarify the appraisal problems.

When models (such as those used for short-term economic forecasting) produce unconditional forecasts, their accuracy can be evaluated straightforwardly by measuring the difference between the forecasted and actual results. However, when models are set up to generate conditional forecasts, their performance can be evaluated in two quite different ways. First, the projection indicated as most likely may be considered as if it were the unconditional forecast nestled among other conceivable but less likely possibilities. This most likely projection may then be evaluated like a standard unconditional forecast: retrospectively, by measuring the discrepancy of predicted trend from actual trend; for current forecasting efforts, by examining the spread of "most likely" projections from several sources.

However, in recent years, particularly since the 1973–74 jump in oil prices, energy forecasters have been far less willing to presume that a particular energy-supply condition, or policy-choice scenario, is more likely than the others. The set of conditional forecasts (or "policy simulations") is offered as an aid to policy-makers rather than as a prediction of what they will do (Congressional Research Service, 1976). No single projection can be meaningfully regarded as the forecaster's prediction.

When no scenario is designated as most likely, the scenarios must be regarded as exogenous factors, whose likelihoods are not at issue in the modeling exercise. The model produces a set of projections, each posited as correct if the corresponding condition or scenario were to hold, but without implying that any particular one will hold or that some are more likely than others. In this case, the retrospective evaluation of forecast accuracy must proceed by first establishing which condition actually prevailed, and then measure the discrepancy between the projection tied to that condition and the actual level of the predicted trend. If it is still too early to evaluate a set of conditional forecasts retrospectively, the spread of conditional forecasts of the same trend for the same target year can be used as one indication of uncertainty or minimum error, but only if the conditional is the same for every forecast of the set. For example, one model's prediction of petroleum demand in a target year t_{i+N} under scenario S_1 can be compared only with the predictions of other models for that year t_{i+N} and that scenario S_1. The differences among the petroleum-demand predictions for t_{i+N} necessarily reflect error, since only one (or none) of them can be correct.

This last approach is the only feasible one for judging the performance of complex energy models. They are too recent for their accuracy to be judged retrospectively, and they project several different conditionals without designating one as most likely. Because it makes no sense to compare the projections of several models if they take different scenarios as their conditional elements, many opportunities for comparison and measurement of dispersion are precluded. When modelers work independently, there is no reason to expect that any of the scenarios examined will be precisely the same for two or more of them. One cannot say, then, whether differences in projections reflect these scenario differences rather than indicating disagreement, uncertainty, or error.

Comparability is feasible, however, when modelers work together to examine the implications of separate models run with a common set of scenarios. The most comprehensive and careful effort in this vein was conducted by the Modeling Resource Group of the National Research Council's Committee on Nuclear and Alternative Energy Systems, with the participation of four modeling groups projecting energy consumption and prices to the year 2010 and beyond, and with limited participation of two medium-range models [7]. Each of the long-term models projected the outcomes of six different policy scenarios, using the same starting points and common assumptions regarding economic growth. Therefore differences in outcomes from one model to another, for the same scenario, represent at least the minimum degree of uncertainty or error of the models.

There, are, in fact, some serious divergences between different models' results. In forecasting energy consumption, the base case projections for the year 1990 vary up to 19% for total energy, and by more than 50% for electricity generation. Some models project five times the consumption of oil than other models; even when the two extreme projections (of the five models involved in this particular comparison) are dropped, one remaining projection of oil consumption is less than half the levels forecasted by the other two models. One model projects more than twice the consumption of nuclear energy than the others (National Research Council, 1978, p. 47).

Most seriously – because the rationale of such simulation exercises is to determine which policies produce optimal results – shifts from one policy scenario to another bring different changes according to different models, especially for the three models most extensively compared in the Modeling Resource Group experiments. For example, the Brookhaven DESOM model and the Stanford University ETA model foresee little impact of a moratorium on nuclear development and limits on coal and shale oil exploration (compared to the base case), while the Nordhaus model (developed by William Nordhaus of Yale University) projects a 20% decline in energy consumption and domestic production under the moratorium conditions.

For energy price projections, there are also major discrepancies. While the ETA and DESOM models project only small differences among different scenarios for the 1990 price of oil, the Nordhaus model forsees a 60–70% greater price under one scenario (viz., limits on coal and shale oil production). Oil prices for the year 2010 forecasted by the ETA and DESOM models are consistently a third to a half greater than those projected by the Nordhaus model. Under all of the scenarios involving constraints on coal and shale oil development, the price of coal is projected to increase two-fold or more between 1990 and 2010 according to the ETA and DESOM models, yet the Nordhaus model projects the price to decline by nearly half. For electricity prices, ETA projects sharp increases from 1990 to 2010 under the scenarios of nuclear moratorium and limits on coal and shale oil production; Nordhaus projects declining prices (National Research Council, p.48).

Although there are numerous consistencies across models, there are also many inconsistencies. The unavoidable conclusion is that some of the models must be wrong in quite significant ways on policy-relevant issues. It is also discouraging that even the agreement across models need not be an indication of validity; they could all be wrong. For example, all energy models predicting the 1975 levels of U.S. electricity, petroleum, and total energy consumption projected these levels higher than they actually turned out to be (Ascher, Ch. 5). This confident consensus was no guarantee that the models were correct then; any consensus among models' predictions in the future may be equally misleading.

3. The Nature of Complex Modeling

The key attributes of complex modeling are explicitness and complexity. While complex modeling has no monopoly on either, the combination is unique. Explicitness and complexity are the virtues of complex models – if and only if the models are valid. When models are misspecified, their explicitness forecloses the possibility of adjusting their results on the basis of judgment and plausibility; the model operator must do "what the model says." Complexity becomes a burden rather than a virtue for a misspecified model because complexity makes it difficult to determine just where the errors are.

3.1. Explicitness

Complex models are formulated by specifying assumptions and hypothesized relationships as explicit, usually mathematical propositions. While this procedure is often very helpful in uncovering inconsistency and vagueness in the initial ideas or verbal formulations, it cannot establish the correctness of the model's propositions. Models express assumptions, but do not validate them. If the modeler tries to ensure the validity of the model's propositions by focusing on disaggregated behavior of presumably greater regularity, the problem of reaggregating these behaviors to model overall patterns becomes another potential source of error. If the modeler only includes relationships proven by past experience, there is no guarantee they will hold in the future. There is no procedure or format of model specification that guarantees the validity of this specification.

With explicitness and the computational capacity of computers, complex models can combine almost immediate response with the fundamental analysis that an elaborate model – if correct – provides. Other explicit mechanical modes of analysis and forecasting, ranging from trend extrapolation to multiple regression, may be just as quick and automatic, but cannot make as bold a claim to represent fundamental understanding of the dynamics of the system under study [8]. If the modeler wishes to convey a detailed image of reality through his model, explicitness will require the model to be large and elaborate. Whereas the analyst relying on judgment keeps the richness of detail and nuance stored in his head, the modeler must commit it all to formal expression. Greater elaborateness has been the most striking trend in model building in the past fifteen years.

Yet explicitness also exercises a subtle constraint on the model's capacity to fully represent reality. Explicitness tends to limit the amount of context encompassed by the model because the need to formulate explicit, consistent relationships requires the modeler to discard those factors which are of some relevance but without a clear-cut, consistent relationship to the outcomes. In modeling approaches there is nothing of the notion of a "directed search" to identify contextual factors which may be relevant but only in particular cases or in idiosyncratic ways. In short, modeling, like extrapolation and other routinized, explicit procedures, tends to simplify and restrict contextual considerations.

The fully explicit model, once formulated, can be applied mechanically, independent of any further judgment on the part of the modeler. Although the model represents the model-builders' assumptions about how the system operates, mechanical operation can thereafter insulate the results from both the modeler's myopia and from his insights about the specific situations being forecast. If the model truly embodies a fundamental understanding of reality with greater sophistication than the casual or less systematically formulated opinion of the model operator, its mechanical operation is a virtue. There are numerous documented instances of analysts incorrectly disregarding the signals of their methods because of personal biases and faulty preconceptions. Yet if the model produces silly results, the lack of intervention by its operators would be a deficiency. Model operators – especially when they are also model builders – are often experts in the substantive area of the model; hence their judgment might be an important means of screening model results. But this alternative of "plausibility testing" is correspondingly problematical, as we shall explore below.

3.2. Complexity

The "complexity" of complex forecasting models lies in the existence of several (and often many) "interdependent" relationships. If some variables (or factors) appear in more than one equation (or proposition), the relationships are "interactive," producing overall patterns of outcomes which would not be obvious from the isolated consideration of each relationship [9].

This sort of complexity should be distinguished from the elaborateness of the accounting devices included in many models (both complex and not) to recombine disaggregated trends or to keep track of their changes over time. Very often these devices appear to be the most impressive methodological parts of a model; in fact they do not represent the theoretical complexity of the model nor the behavioral complexity of the phenomena being modeled. For example, an energy model with fifty different equations for the consumption levels of fifty different fuels, and a few more equations to add up consumption for broader fuel categories, may nonetheless be limited to quite simplistic propositions for each specific trend; e.g., that the consumption of a given fuel increases by a fixed proportion each year. In contrast, a model linking the consumption level of one broadly-defined fuel category to the levels of a few other broad fuel types (e.g., petroleum consumption limited by a large increase in coal consumption, which in turn is stimulated by lagging nuclear energy generation) may be considerably more complex.

A set of interactive relationships has emergent properties, often not apparent to the analyst who examines these relationships one at a time without an explicit model to track the interactions. Though the emergent patterns are indeed nothing more than implications of the basic relationships and the explicitly expressed connections among them, these patterns nonetheless can be surprising, counterintuitive, and inconsistent with expectations.

The desirability of such surprising outcomes from a model is one of the few points of disagreement among modelers in terms of the objectives of modeling. On the one side, Jay Forrester has stressed that the advantage of complex models is precisely in their ability to establish the counterintuitive implications of their component propositions (Forrester, 1968; Forrester, Mass and Ryan, 1976). If the implications of one's assumptions and hypotheses are obvious when they are "thought out" without the aid of a model, why use one? If the implications are surprising, the strength of the complex model is in drawing from a set of assumptions these implications that otherwise would be discarded because they are inconsistent with preconceptions. The mechanical operation of a computer model, unencumbered by plausibility testing or modifications by their operators to bring the model's outcomes into congruence with their own conceptions, in effect insulates the task of "implication-searching" from the influence of preconceptions.

4. The Nature of Forecasting

When complex models are applied to the task of forecasting, they enter a field with a track record antedating complex modeling. Thus, although modeling itself is a recent addition to the tools of forecasting, some insights can be drawn both from a theoretical examination of what forecasting entails, and from the practical benefits and caveats of different forecasting approaches.

4.1. Theoretical Considerations

Philosophers of science point out that there is little formal difference between prediction and retrospective explanation [10]. An initial array of "descriptors" of past or present situations is linked to another array representing outcomes, through a set of procedures comprising the predictive or explanatory schema. This schema may require three additional kinds of information beyond the initial array (which is sometimes called the "state vector"):

(1) Conditionals: facts establishing which propositions, routines, or procedures of the schema are to be invoked. Since many schemas include conditional propositions (e.g., "If X occurs, then Y will follow, or relationship Z will hold"), the conditionals must be provided to "drive" such schemas.
(2) Parameters: quantitative factors treated as fixed for any specific application of the schema, but which can take on other values for other applications of the same schema. Parameters are usually embedded within the schema's propositions or procedures, but, in contrast with "variables", are not determined by the schema nor usually through the logic and assumptions underlying the schema. Thus they too must be estimated apart from the operation of the schema.
(3) Exogenous variables: variables utilized in the schema but not determined by it.

Consider, for the sake of illustration, a simplistic schema to account for the overall level of U.S. defense expenditures. The schema consists of two propositions: i) In a period of "detente," the defense budget of the U.S. increases at a fixed annual rate; ii) In a period of "confrontation," the U.S. defense budget increases at a fixed proportion of the rate of the Soviet Union's budget increase of the previous year.

For this schema, the initial array consists of the U.S. and the Soviet defense budget levels of the years just prior to the first year the schema is to cover. The conditional is detente or confrontation (hopefully clearly and dichotomously defined); the parameters are the rate of increase involved in the first proposition and the fixed proportion found in the second; the exogenous variables are the Soviet budget levels.

The practical difference between explanation and prediction is that for retrospective explanation (e.g., "Why did the U.S. defense budget go up as rapidly as it did?"), the conditionals, parameters and exogenous variables generally can be supplied with a high degree of certainty (since they are already known, within the limitations of measurement), while in prediction (e.g., "What will the U.S. defense budget be at some future date?"), they are themselves matters for forecasting. For prediction, parameters cannot be estimated on the basis of past patterns with any a priori assurance that these patterns will continue into the future; parameters must be projected rather than estimated. For prediction, the conditionals and exogenous variables cannot be found in the knowledge of what has already occurred, but rather must be forecasted outside of the original schema. Thus, in practice the nature of information available for predictive and explanatory efforts differs: for explanation the information is drawn from existing data; for prediction it must be projected through judgment or more mechanistic techniques. Even the nature of the schema may differ in explaining and predicting the same phenomena, because the predictive schema will be designed to avoid information which is difficult to project even if it is superb for explanation. For example, an explanatory schema designed to account for historical trends in income distribution may rely on known changes in governmental wage policy, while a forecaster may be loathe to predict policy and instead may focus on productivity trends.

These considerations imply, first, that prediction is a much more formidable task than explanation, and second (a much less obvious point), the qualities of good predictive models are not necessarily the qualities of good explanatory models. Because some information taken for granted for explanation remains unknown (and outside of the schema) for prediction, forecasting is subject to one important source of error that is avoided by retrospective analysis: the occurrence of events or conditions which are unanticipated or regarded as unlikely before the fact, but of obvious importance once they materialize. Any reasonably complete explanatory model of world economic patterns in the twentieth century would certainly take into account the impact of the First World War, yet predicting the occurrence and impact of the war prior to 1919, or basing a projection on the presumption that it would occur, would have been a very stiff challenge for prediction. Therefore, in practical terms, forecasting requires a much broader consideration of the wide range of possibly pertinent factors than does explanation, simply because such factors cannot be included or excluded a priori. For example, in 1972 an economist formulating a model to account for the historical pattern of the American economy could legitimately dispense with modeling the mechanism of drastic oil-price increases imposed by a petroleum cartel, as well as the economic impact of energy costs as a newly prominent limiting factor. Testing this model with historical time series data very well could prove it accurate even if it had ignored OPEC, yet as a forecasting model applied beyond 1972 it would have been seriously deficient.

The parameters found that fit in the past will hold in the future only if the relationships embodied by the model are invariant. Simply finding a set of relationships and parameters that fit past data does not guarantee future fit. Some theorists argue persuasively that if such relationships describe aggregate behavior (e.g., macroeconomic behavior in the case of economic models), invariance is unlikely even in principle. Aggregate specifications are subject to "parameter drift," if only because individuals' decision rules change from one situation to another, from one set of expectations to another (Lucas, Jr., 1976). Robert Lucas points out that even the parameters utilized in the large scale econometric models as if they had been constant have in fact been drifting, and that this is actually acknowledged by the forecasters through their adjustments of model forecasts on the basis of prior model errors (Lucas, Jr., pp.23–24).

We come back to the puzzle of why the large-scale econometric models for short-term economic forecasting are under continual development, and yet do not improve in accuracy. This line of reasoning suggests an explanation: the "improvement" of substituting new specifications for old ones merely replaces the representation of the previous context with a representation of the newer context, but without coming closer to a more generally valid representation. There is no a priori reason why there should be a set of aggregate-level propositions that are generally valid over time.

4.2. Considerations from Retrospective Appraisal

A retrospective appraisal of the forecasting record provides more specific insights. The record compiled to evaluate the accuracy and biases of various forecasting methods and sources consists of all available forecasts of specific U.S. national trends in population, energy demand (viz., consumption of petroleum, electricity and total energy), transportation (viz., airline passenger: volume, general aviation fleet size, motor vehicle registrations), economic growth (viz., real and nominal GNP), and technology (viz., computer-speed capability and nuclear energy capacity (Ascher, 1978). Reviewing the forecasting record leads to three general conclusions.

First, methodological sophistication contributes very little to the accuracy of forecasts. The introduction of more sophisticated methods in population forecasting, with the elaborate accounting divisions of components and cohorts, has not resulted in more accurate demographic forecasts. We have seen that econometric modeling has not improved forecast accuracy. Correlation and simple trend extrapolation have had the same general levels of accuracy in predicting energy-demand trends, even though correlation is a much more elaborate procedure. Separating electricity or petroleum demand into various end-use components does not add any appreciable accuracy over projecting the demand for each as a unitary trend. In transportation forecasting, the more elaborate regression models recently adopted by the FAA to predict both commercial air traffic and the general aviation fleet size have not improved the accuracy record.

Second, a forecast's time horizon is the strongest and most consistent correlate of its accuracy. Though there are some exceptions, the general rule is that shorter forecasts are more accurate, often in a nearly linear relationship. For example, 5-year petroleum consumption forecasts have had a median error of about 6% and 10-year forecasts an error of about 13%; motor vehicle registration forecasts have median errors in percentages roughly equal to the forecast lengths in years. A direct implication of the importance of forecast length is that reliance on elaborate, previously made forecasts, either for direct use or as a basis for projecting related trends, can be very costly if it entails the use of what are, in effect, longer forecasts. This is the real cost of the one-shot definitive efforts used long after they were produced, and also of utilizing already published – and hence to some extent out-of-date – projections.

Third, all evidence points to the essential importance of the validity of core assumptions antecedent to the choice and application of methodology. Behind any forecast, regardless of the sophistication of methodology, are irreducible assumptions representing the forecaster's basic outlook on the context within which the specific trend develops. For example, envelope curves are chosen as a forecasting method only if the forecaster has a preconception that cumulative scientific breakthroughs will cause the technology improvement rate to take the envelope-curve form. Forecasting by analogy is utilized only when the forecaster believes a particular historical pattern to be analogous. Formal models, too, ought to be viewed as the explicit expression of a set of assumptions, and the operation of the model as the tracking interactions arising from these presumed relationships. Core assumptions are not derivable from methodology; on the contrary, methodology represents the vehicle for tracing through the consequences or implications of core assumptions originally chosen independently of (and intellectually prior to) the method; it reflects and signifies this preconception of the pattern of future change.

These core assumptions are the major determinants of forecast accuracy. When the core assumptions are valid, the choice of methodology is either secondary or obvious. When these assumptions fail to capture the reality of the context, other factors such as methodology generally make little difference: they cannot redeem a forecast based on faculty core assumptions. This helps to explain why correlational forecasts of electricity and petroleum demand do not perform better than extrapolation; when one trend is projected by correlating it with a presumably more fundamental, contextual trend (say, industrial production or GNP), its accuracy depends not primarily on whether this correlation is appropriate, but rather on how well the more fundamental trend is forecast. The primacy on contextual assumptions is also consistent with the great dispersion of projections of air transportation forecasting models based on regression of economic growth; the greatest uncertainty is in predicting the economic growth context (Ascher, p. 164).

The importance of both recency and core assumptions makes the problem of relying on antiquated core assumptions particularly serious. This "assumption drag" has been the source of the most drastic errors in forecasting. The worst population forecasts prepared in the late 1930s and early 1940s were based on a no longer valid assumption of declining birth rates. Similarly, the electricity-demand forecasts of the early 1960s continued to project fairly low electricity-demand growth, even when the actual growth rates contradicted this assumption.

What are these core assumptions whose validity is so crucial? Often, they are political. Even for technological forecasting, seemingly far removed from political developments, a major source of uncertainty stems from doubt about the political context. If the variation among predictions of a given technological breakthrough is taken as a measure of uncertainty, it can be demonstrated that developments in technological areas where advances require large-scale official programs (e.g., health care systems, medical education, space exploration) are more uncertain than developments in technologies depending on engineering refinements and the disaggregated market diffusion of such innovation (e.g., communications, educational technology, automation) (Ascher, pp. 190–191). The difficulty in predicting innovations requiring discrete high-level "official" (though not necessarily governmental) policy decisions indicates the pivotal role of political assumptions – both because the political context is important to the development of these technologies and because it is difficult to predict. Political forecasting, perhaps because of the discrete and discretionary nature of political decisions, is the weakest link in the whole range of forecasting trends even remotely affected by policy. The political context is often ignored (on the dubious grounds that in the long run, relatively short-lived political conditions or policy decisions will have no net effect), or it is considered in a rudimentary fashion. For example, almost all energy projections made prior to 1973 totally ignored governmental responses to energy problems, including the possible responses of petroleum-producing nations to the dwindling real earnings of their oil exports. Even now, long-term energy, economic, and transportation forecasts, if they are based on any explicit assumptions about government policy at all, generally pose a single and stable governmental policy choice, exogenous to the forecasting procedure, as a fixed condition (McDaniel, 1978; Congressional Research Service, 1976; and Dewhurst and Associates, 1947).

This becomes a serious limitation because it overlooks the possibilities of problem-solving or equilibrating actions of government response to the trends as they unfold. For example, if a constant or consistently increasing level of governmental conservation is presumed for the entire forecast period, the projection would be insensitive to the possibility that if energy consumption nevertheless becomes excessive, the government might redouble its conservation efforts.

4.3. Implications for Modeling

If, indeed, valid up-to-date contextual assumptions, rather than methodological sophistication, have been the primary determinants of forecast accuracy, there is little basis for confidence that the unique capacity of models to explicitly track complex interactions will materialize in consistently greater accuracy. The advantages of complexity and explicitness may well be offset by the built-in assumption drag of elaborate models, as well as by their tendency to restrict contextual considerations to regular, law-like relationships.

Just as importantly, complex modeling, though it certainly does not preclude the incorporation of policy interactions, has failed to do so. The current generation of complex models developed for economic and energy modeling do incorporate socio-political factors, including policy choices, but only as givens, usually expressed through conditional scenarios or parameters. Complex modeling has been particularly slow to incorporate policy response. This omission is illustrated by the approach of William Hogan (organizer of the Energy Modeling Forum at the Stanford University Institute for Energy Studies) and Alan Manne (a principal developer of the Stanford ETA energy model) in an article on economy-energy interactions. They argue that:

> The value share of the energy sector determines the incremental effect upon the GNP. If the 4% value share remained constant, this would mean that a 10% reduction in energy inputs would produce only a 0.4% drop in total output. Thus, for small changes in energy availability, there need not be a proportional impact upon the economy as a whole (Hogan and Manne, 1977).

They go on to say that for large changes in energy availability, the critical factor in determining economic growth rates becomes the substitutability of energy inputs by non-energy inputs (e.g., insulation, energy-efficient machinery, etc.) to compensate for energy supply shortfalls (Hogan and Manne, pp. 249–250). This formulation may be reasonable if the government is passive or acts only to facilitate the substitution of inputs. But it does not encompass the possibility that the government may react to inflation (triggered by higher energy prices) by enacting recessionary economic policies with much greater impact on economic growth. The commonality of recessionary policies adopted by OECD countries after the oil price increase of 1973–74 would indicate that this sort of reaction is regular enough to be modeled with some degree of confidence.

The failure to model policy response is, of course, not a mere oversight. It reflects the almost universal commitment of practical modelers to represent aggregated dynamics with aggregate, "macro" models, in which discrete decisions (if not provided as givens) are assumed to be subsumable by general mechanisms of equilibration or optimization.

In principle, disaggregated models based on decision-making agents as the units of analysis (e.g., based on the perhaps firmer principles of microeconomics) can model both discrete responses and the aggregate patterns. If indeed the aggregate behavior of the system follows a pattern of equilibration or optimization as a result of myriad individual decisions – a point not challenged by macroeconomics – a micro-level model can represent such patterns.

5. The Practical Application of Modeling to Forecasting

The formal properties of forecasting and of complex modeling do not completely determine how complex models are applied in practice to forecasting tasks. There is a range of discretion in the practice of forecast modeling. Choices within this range can make a huge difference in the utility of complex models.

Despite the large number of possible model forms, varying in size, format, level of aggregation, propositional form, etc., in practice complex forecasting models have become highly uniform. Once practical forecasting in a particular area becomes established, there is a strong tendency for the models to converge. Those showing the earliest success drive out the less successful. From the interesting variety of econometric forecasting models developed in the past, the survivors are remarkably similar: all are large, macroeconomic models in the Klein-Goldberger tradition [11]. This could be justified as "survival of the fittest," except that complex modeling as an intellectual enterprise is not advanced enough for all other lines of development to be rejected as inferior simply on the basis of early falterings. In particular, micro-level models have received surprisingly little support and development, considering their initial promise [12].

Thus far, the application of complex modeling to forecasting has also been marked by the related practices of commitment to the same basic structure for long periods of time, continual superficial revision of such models and plausibility testing of their outputs. Most methodologically sophisticated forecasting efforts are big, expensive, one-shot affairs. Since rigorous, elaborate analysis is time consuming and expensive, there has been a natural tendency for forecasters to pour their efforts into grand, once-and-for-all projects, carried out only infrequently and yet used long after they are produced because the immense effort makes them seem definitive. This of course, poses the problem of obsolescence.

In one important way, it would seem that modeling approaches depart from this pattern of one-shot projects because they are automated, ongoing mechanisms which can be updated, and which can spew out "new" forecasts at any time. Yet the sunk costs in the development of an elaborate model forces it to be a one-shot effort in a more basic sense. Though forecasters may tinker with their model and resestimate its paramaters, the models tend to remain intact for long periods of time. It is much easier for the modeler to tinker than to scrap the entire model, even though the core assumptions may reside in the model's basic structure. For example, the central mechanism of some energy models is an optimization routine that establishes supplies, demands and prices so as to maximize benefits under the presumption of competitive markets (National Research Council, p.5). Charles Hitch points out that such models (among others) "assume that the economy and its markets wil be permitted to function, that prices are not controlled, and that producers and consumers will not be constrained by regulation from responding to appropriate market signals," (Hitch, 1977, p.iv). Yet, after the modeler has spent years developing optimization routines, apparent violations of such assumptions are more likely to be accommodated by patchwork modifications, or disregarded altogether as short-term aberrations, than they are to trigger the abandonment of the model altogether.

Even though these basic structures endure, complex models in practice undergo continual revision. Although the primary emphasis of forecast modeling is usually on the quality of the forecasts rather than on the development of the models, forecasting models (like complex models in general) still represent efforts at theory building and this effort is never finished. Therefore model revision, which seems to the cynic to be an ad hoc effort to keep a fundamentally misspecified model more-or-less in line with reality, is often regarded by the model builder as the normal routine of science. Continual revision complements rather than contradicts the durability of basic model structure because altering the same basic model to keep it consistent with the most recent changes in the phenomena it models saves the modelers from scrapping the basic model itself.

Plausibility testing is an almost universal procedure in the practical application of forecasting models, resulting in part from the commitment to continual model development and in part from the modelers' modesty as to the accuracy of their models. Usually, at least at this stage in the development of modeling, when a model generates implausible outcomes the onus is laid to the model rather than to the modeler's perception of what is plausible. Thus plausibility testing proceeds not just by directly rejecting implausible model results, but also through the subtler means of providing exogenous information that makes the model's results plausible. The econometric modelers come up with plausible results by:

> changing the preliminary assumptions on exogenous variables and constants until the resulting forecast falls within the range thought to be reasonable – which is principally responsible for the improvement in ex ante forecasts. From this point of view, the model serves primarily to assess judgmentally the general implications of the forecaster's assumptions on future exogenous developments, including his ad hoc adjustments for anticipated changes in structure since the sample period, and for the correction of apparent specification errors (Hickman, 1972, p. 17).

In energy modeling, a good example of the interplay between exogenous value selection and the plausibility of model results is provided by the treatment of energy prices in the RFF/SEAS Modeling System:

> Since the model does not include a mechanism for determining these prices, our procedure is to develop assumptions about price changes, run the model, adjust prices, and iterate if neccessary (Ridker et al., 1977, p. 141).

What are the combined effects of these practices? The complex model holds an ambiguous status as both practical device and experiment. When model builders come up with new formulations, should forecast users regard such models as the best devices the modelers can offer for forecasting, or as the latest turn, perhaps a dead end, in the theoretical and methodological search? Most importantly, there is an obvious and direct trade-off between plausibility testing and the capacity of the model to express the counterintuitive implications of its assumptions. These practices drastically weaken the forecast modeling enterprise's sensitivity to surprising future possibilities. The homogeneity of models simply reinforces this weakness, as similar models undergoing similar judgmental censorship by modelers holding similar outlooks on the future can so easily reassure all parties that the future is seen with certainty.

It is not clear, however, how long these practices will last. Certain models have become institutionalized, each with a life and a durability beyond the active role of its originators. Related to this process of institutionalization is a growing distinction between model builders and model operators. Innovators like Lawrence Klein and Otto Eckstein are not preoccupied with the routine operation of the Wharton and DRI econometric models; they have other, more ambitious things to do.

This development presents both drawbacks and advantages. When methods themselves acquire charisma, their operation can take the place of careful reasoning. Model operators who regard themselves as technicians are even less likely than model builders to grapple with the basic structure of an already established model. Hence the obsolescence of core assumptions embodied in the model's structure will become an even greater risk. If any judgmental input remains in the operation of the model, it will not be from leading economists privy to the discussions at the highest levels of government and industry. Thus, even if the model operators are no less competent than prominent model builders as practical economists, the caliber of "inside information" used in the remaining practice of plausibility testing is likely to deteriorate.

The advantage of this development of institutionalized models is in permitting the models to engage in the admittedly risky enterprise of forecasting free of further plausibility testing. This does not mean that complex models would be, or should be. freed of all constraints on plausibility; these still reside in the relationships constituting the model. If there are counterintuitive surprises on the horizon, and if the models are really good enough to anticipate them, then the unhampered model will serve a vital purpose.

6. Conclusions

The issue boils down to this: if complex models as normally operated converge with judgment, or, if operated more mechanically, perform no better than judgment, what is the incentive to employ an expensive, formidable model? The current practice of model operation, which is virtually "judgment guided by modeling", is saddled with the predictive weaknesses of aggregate macro-level models, without the capacity to reveal surprising outcomes implicit in the component propositions. The possibly emerging practice of mechanically-run aggregate models could restore this "sensitivity to surprise," but falls prey to the most severe problems of "assumption drag," inadequate representation of policy response, and parameter drift.

This dilemma is irresolvable as long as all complex forecasting models are basically alike and are operated alike to pursue the same function. More pluralism in practical forecast modeling is urgently needed. First, the two functions of forecast modeling, to produce consistently reasonable forecasts and to anticipate counterintuitive outcomes, can rarely be pursued by the same model at the same time, but as separate enterprises can be mutually reinforcing. The former provides credibility for the latter, while surprise sensitivity, as the unique contribution of modeling, enhances the attractiveness of the entire modeling undertaking. Therefore, the operation of some models in the current mode of plausibility testing, and other models in the surprise-sensitive, unconstrained mode, can benefit both approaches.

Second, the homogenization of practical forecasting models should be avoided; it not only cuts off promising avenues for developing better models, it leaves the field to the large-scale aggregate models which (for all the reasons discussed above) will most likely always require plausibility testing. The really quite modest success of the macro-level modeling operations (i.e., the models plus the judgmentally based interventions of the model operators plus the adjustments for parameter drift) in terms of performance is no basis for concluding that the models per se are optimal in either specific detail or general form.

This diversity in operating modes and model forms requires considerable sophistication on the part of the models' audiences and funders. They must be able to recognize which models' outputs are to be regarded as most consistently and plausibly likely, and which are designed to be ultra-sensitive to possible surprises even at the risk of a worse overall record of accuracy. If this is not recognized, forecast users will be disappointed in the first type for not being daring enough and cynical about the second type on the basis of its overall performance. Funders must be aware of the widely differing development periods necessary to bring different modeling approaches to fruition, and must maintain a broad "R & D" strategy of cultivating different approaches even when only a few seem superior in the short run. Diversity and a modicum of tolerance will aid in bringing out the potential of complex modeling, without the danger of elevating one complex modeling approach as the panacea for the problems of forecasting.

Notes

1. We are excluding models in which overlapping or shared variables could be eliminated through (1971), Essay 2.
2. Except for "identities", i.e., equations that aggregate without representing behavioral properties.
3. See Richard Vitek and Nawal Taneja (1975) as well as Dale E. McDaniel (1978) who reports that after extensive searching, he was able to find only three transportation models that could be termed "complex" in the sense we are employing.
4. "Short Term" in economic forecasting generally means no more than four to six quarters. The survey of these short-term forecasts is published in both *The American Statistician* and *Explorations in Economic Research*, various issues.
5. See Vincent Su (1978). It is important to note that the error of the median forecast is not the same as the median error of the set of forecasts.
6. See the very candid description given by Lawrence Klein (1968).
7. See National Research Council (1978). Two other comparative efforts, reported in Charles Hitch (1977) and William Hogan (1977) did not enforce as much uniformity in scenarios, assumptions and starting points. In any event, there is much overlap in the models examined.
8. See the delightful discussion of the fundamentalist-chartist controversy in Martin Shubik (1967).

9. The issue of emergent properties is presented most cogently in Ronald Brunner and Garry Brewer (1971), Essay 2.
10. This issue is systematically reviewed by Carl G. Hempel (1965), Ch. 12. Hempel speaks of "explanatory and predictive models"; the use of the term "schema" is adopted here to avoid confusion with the term "model" as used throughout his article.
11. See Greenberger et al., (1976); Chapter 6 has a discussion of the genealogy of econometric models.
12. For example, Guy Orcutt, et al. (1961), for an early effort, and George Sadowsky, (1975). Such models have largely been relegated to examinations of local systems or partial analyses of specific policy options (e.g., income maintenance programs). These applications are discussed in Greenberger, et al., (1976), pp. 107–115.

References

Ascher, William (1978). *Forecasting: An Appraisal for Policy-Makers and Planners.* Baltimore: Johns Hopkins University Press.

Brunner, Ronald and Brewer, Garry D. (1971). *Organized Complexity.* New York: The Free Press.

Congressional Research Service (1976). "Energy demand studies – An analysis and comparison," *Middle- and Long-Term Energy Policies and Alternatives,* Part 7, Appendix to Hearings Before the Subcommittee on Energy and Power, Committee on Interstate and Foreign Commerce, U.S. House of Representatives, March 25–26, Washington, D.C.: U.S. Government Printing Office.

Dewhurst, Fred and Associates (1947). *America's Needs and Resources.* New York: Twentieth Century Fund.

Forrester, Jay W. (1968). *Principles of Systems.* Cambridge, MA: Wright-Allen Press.

Forrester, Jay W., Mass, Nathaniel and Ryan, Charles (1976). "The system dynamics national model: understanding socio-economic behavior and policy alternatives," *Technological Forecasting and Social Change* 9 (July): 51–68.

Greenberger, Martin, Crenson, Matthew and Crissey, Brian (1976). *Models in the Policy Process: Public Decision Making in the Computer Era.* New York: Russell Sage Foundation.

Hempel, Carl G. (1965). *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science.* New York: Free Press.

Hickman, Bert G. (1972). "Introduction and Summary," in Bert G. Hickman (ed.), *Econometric Models of Cyclical Behavior.* New York: National Bureau of Economic Research.

Hitch, Charles (ed.) (1977). *Modeling Energy-Economy Interactions: Five Approaches.* Research Paper R-5. Washington, D.C.: Resources for the Future.

Hogan, William W. (1977). "Report on the Energy Modeling Forum," Working Paper. Stanford, Cal.: Stanford University.

Hogan, William W. and Manne, Alan S. (1977). "Energy-economy interactions: The fable of the elephant and the rabbit," in Charles Hitch (ed.), *Modeling Energy-Economy Interactions: Five Approaches.* Research Paper R-5, Washington, D.C.: Resources for the Future.

Klein, Lawrence (1968). *An Essay on the Theory of Economic Prediction.* Helsinki: Jahnsson Lectures.

Lucas, Robert E., Jr., (1976). "Econometric policy evaluation: a critique," in Karl Brunner and Allan H. Meltzer (eds.), *The Phillips Curve and Labor Markets.* Amsterdam: North-Holland.

McDaniel, Dale E. (1978). "Transportation forecasting: a review," *Technological Long-Range Forecasting: From Crystal Ball to Computer.* New York: John Wiley.

McNees, Stephen (1976). "An evaluation of economic forecasts: extension and update." *New England Economic Review* (Sept./Oct.): 30–44.

McNees, Stephen (1977). "An assessment of the Council of Economic Advisers' forecast of 1977," *New England Economic Review* (March/April): 3–7.

National Research Council, Committee on Nuclear and Alternative Energy Systems, Synthesis Panel, Modeling Resource Group. (1978). *Energy Modeling for an Uncertain Future.* Supporting Paper 2. Washington, D.C.: National Academy of Sciences.

Orcutt, Guy, Greenberger, Martin, Korbel, John and Rivlin, Alice (1961). *Microanalysis of Socioeconomic Systems: A Simulation Study.* New York: Harper and Row.

Ridker, Ronald et al. (1977). "Economic, energy and environmental consequences of alternative energy regimes. An application of the RFF/SEAS modeling system," in Charles Hitch (ed.), *Modeling Energy-Economy Interactions: Five Approaches*. Research Paper R-5. Washington, D.C.: Resources for the Future.

Sadowsky, George (1975). "MASH: A Computer System for Policy Exploration." Working Paper 5096. Washington, D.C.: The Urban Institute.

Shubik, Martin (1967). "Comments on Working Session Two: The nature and limitations of forecasting," *Daedalus* 96 (Summer): 945.

Su, Vincent (1978). "An error analysis of econometric and noneconometric forecasts," *Proceedings of the American Economic Association* 68(May): 306–312.

Su, Vincent and Su, Josephine (1975). "An evaluation of ASA/NBER business outlook survey forecasts," *Explorations in Recent Research* 2(Fall): 588–618.

Vitek, Richard and Taneja, Nawal (1975). *The Impact of High Inflation Rates on the Demand for Air Passenger Transportation*. Cambridge, MA: Massachusetts Institute of Technology Flight Transportation Laboratory.

Zarnowitz, Victor (1978). "On the accuracy and properties of recent macroeconomic forecasts," *Proceedings of the American Economic Association* 68 (May): 313–319.